

# RollerCoaster Tycoon: An Analysis of Coaster Construction

Claudia Lyu, Ivana Lin, Lingruo Pan, Megha Koshy

## Data Description

The “RollerCoaster Tycoon Data” is a dataset from Kaggle (<https://www.kaggle.com/datasets/nolanbconaway/rollercoaster-tycoon-rides>) that contains information and metrics about various roller coasters that a user has built on the classic video game “RollerCoaster Tycoon”. The dataset contains 142 rows and 23 columns. The quantitative variables we focus on include excitement, intensity, nausea, max positive G-force, max negative G-force, max lateral G-force, max speed (mph), average speed (mph), duration (seconds), ride length (feet), total time experiencing weightlessness (seconds), highest drop height (feet), number of inversions, and number of drops of each roller coaster. Excitement, intensity, and nausea levels range from 0 to around 10 with no definite maximum. We also explored one categorical variable, custom design. Additionally, we manually added additional categorical variables, such as inversions and max\_lateral\_gs to inform our analysis. The dataset included other categorical variables like roller coaster type and the theme of the roller coaster but there were far too many levels in each of the two categorical variables so we did not end up including them in our analysis.

## Research Questions

Using this dataset, we want to answer three main research questions:

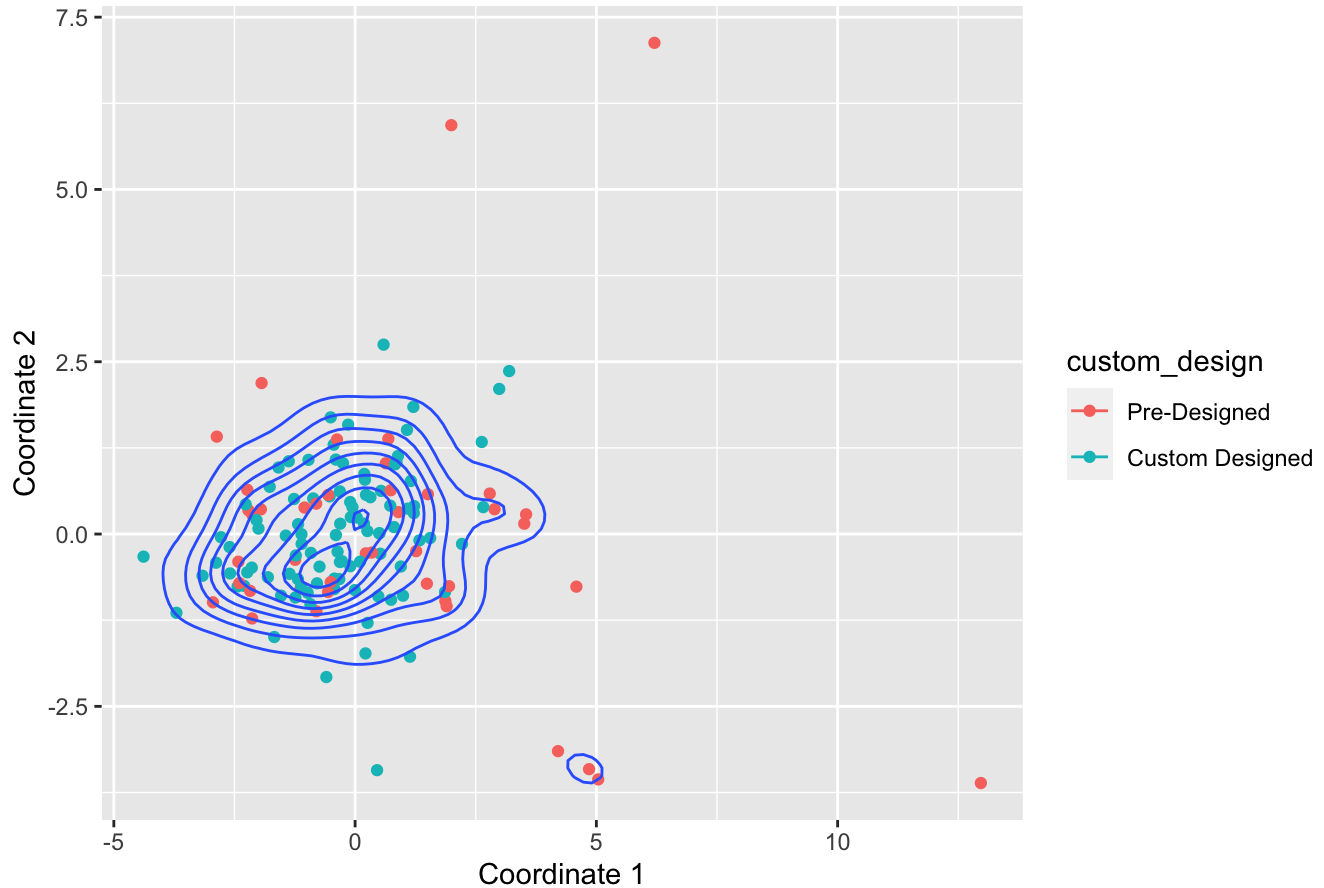
- How do custom design and pre-designed roller coasters similar and how are they different?
- What factors predict the nausea rating of a roller coaster?
- What factors are associated with the excitement rating of a roller coaster?

## Research Question 1: Custom Design vs Pre-designed

To answer the research question, we constructed a multidimensional scaling plot and a dendrogram of all quantitative variables colored by custom design to evaluate if there was any difference between custom designed roller coasters and pre-designed roller coasters. We then plotted a scatterplot of excitement rating vs nausea rating colored by custom design to evaluate how two particular variables of interest may change based on if the roller coaster was custom design or pre-designed.

# Graph 1

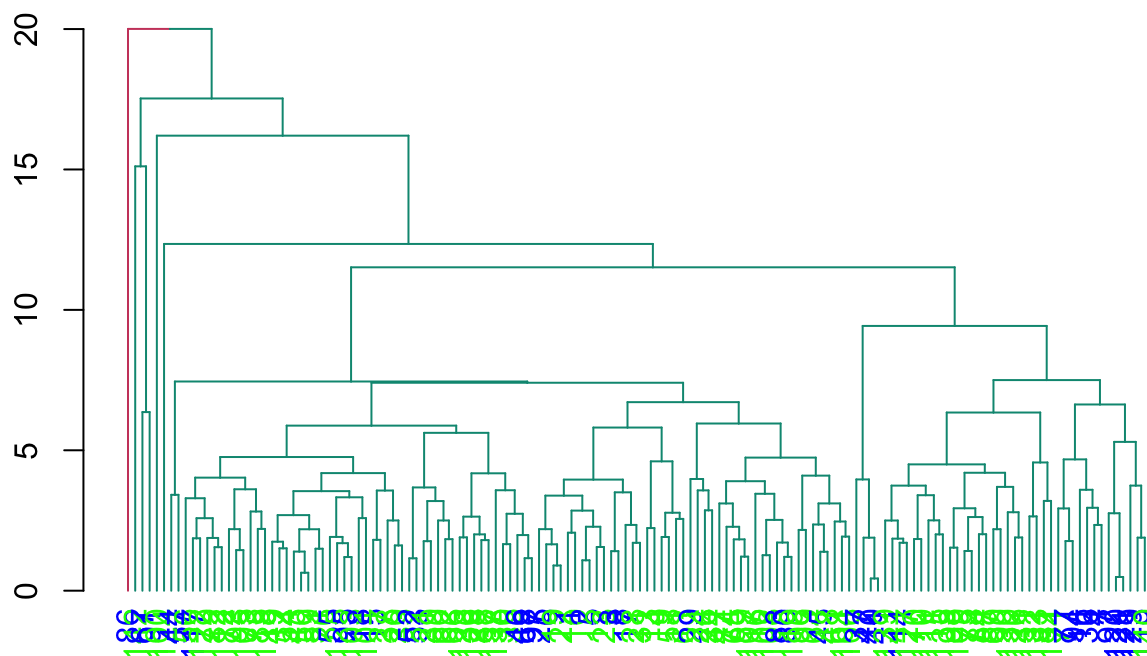
## MDS based plot of Rollercoasters by Custom Design vs Pre-Designed



We see from the plot above that there are two main modes in the left center of the plot and one minor mode in the center bottom of the plot. The two main nodes appear to correspond to custom designed roller coasters while the minor mode looks to correspond to pre-designed roller coasters. Both pre-designed and custom designed roller coasters do not appear to be very tightly clustered, which suggests that there is a lot of heterogeneity in roller coasters that are pre-designed and a lot of heterogeneity in roller coasters that are custom designed. The custom designed and pre-designed roller coasters also overlap quite a bit in their clusters so it seems like there may not be much difference between custom and pre-designed roller coasters.

## Graph 2

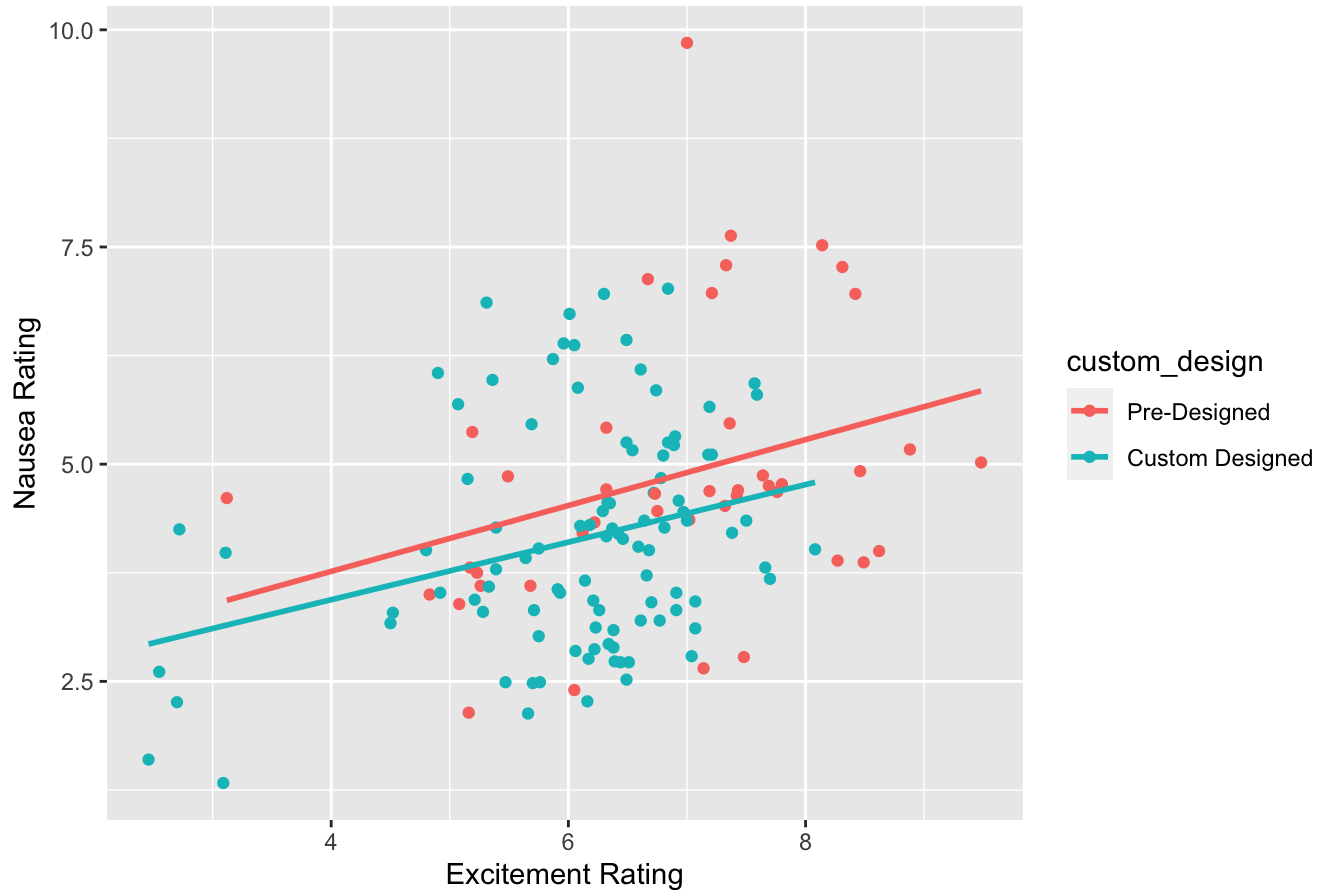
### Complete Linkage Dendrogram of Roller Coasters by Custom Design



The leaves in the dendrogram are colored by custom design of roller coasters. That is, blue leaves correspond to pre-designed roller coasters and green leaves correspond to custom designed roller coasters. There doesn't seem to be any intuitive clustering of the leaves by custom design and the two clusters identified by the dendrogram don't align with the custom design of roller coasters in the dataset. Therefore, we conclude that similar to the MDS plot earlier, there is not much difference between custom designed and pre-designed roller coasters and that custom design and pre-designed roller coaster are similarly distributed.

## Graph 3

Nausea vs Excitement for Custom Designed and Pre-Designed Roller Coasters



Here, we plotted how much the nausea rating increases as the excitement rating increases by custom design vs pre-designed roller coasters. It seems that generally, as excitement levels increase so does the nausea level. However, we see that the increase in nausea for every increase in excitement is slightly lower for custom designed roller coasters than for pre-designed roller coasters as seen through the steeper slope for the pre-designed regression line. It appears that custom designed roller coasters are able to maximize excitement ratings while minimizing nausea ratings better than pre-designed roller coasters since the slope of the line for custom designed is less steep. We can see from the plot that custom designed roller coasters have a higher excitement rating when holding nausea rating constant.

## Research Question 2: Nausea

What qualities do higher nausea levels embody? By taking a look at the correlations between nausea and other attributes, we can understand how various factors relate to different nausea levels.

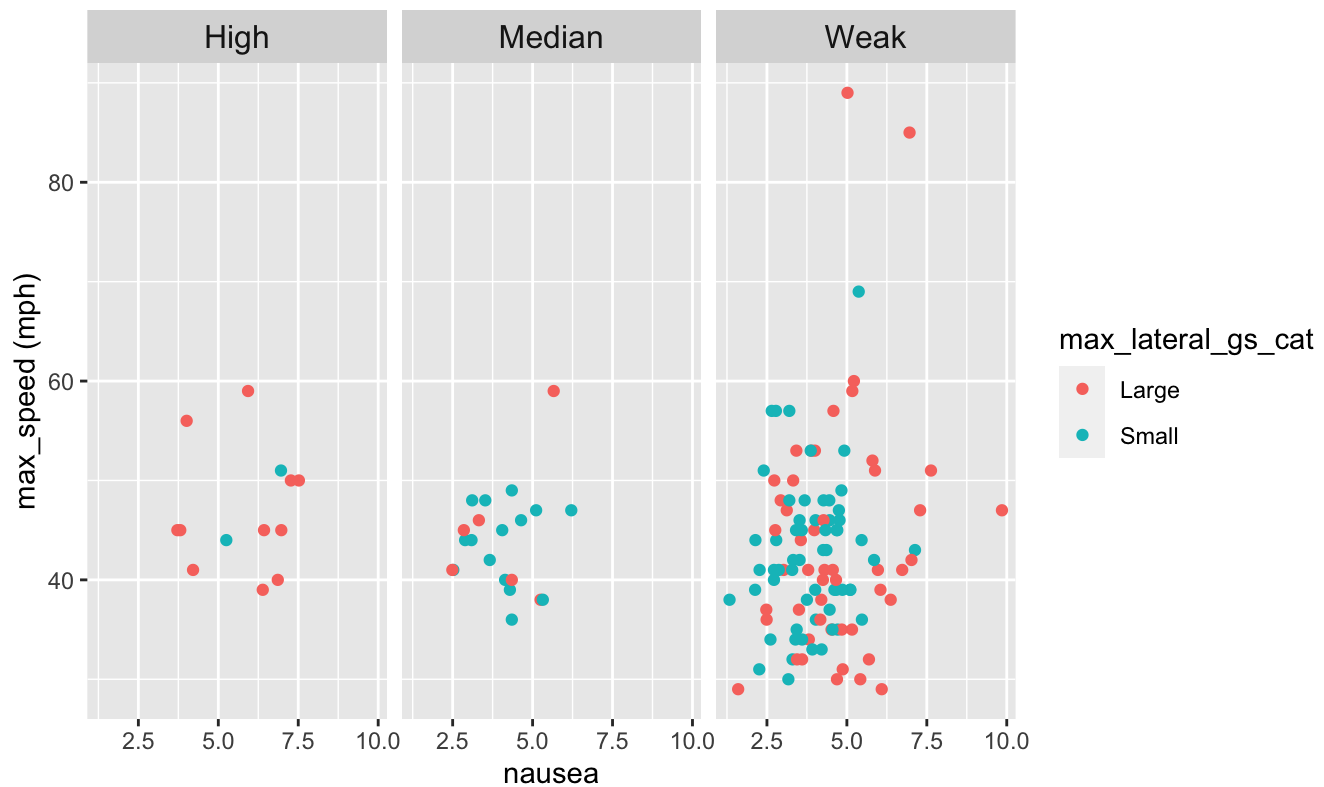
## Graph 4

### Nausea vs Max\_speed, by Inversions and Max\_lateral\_gs

High = high inversions

Median = median inversions

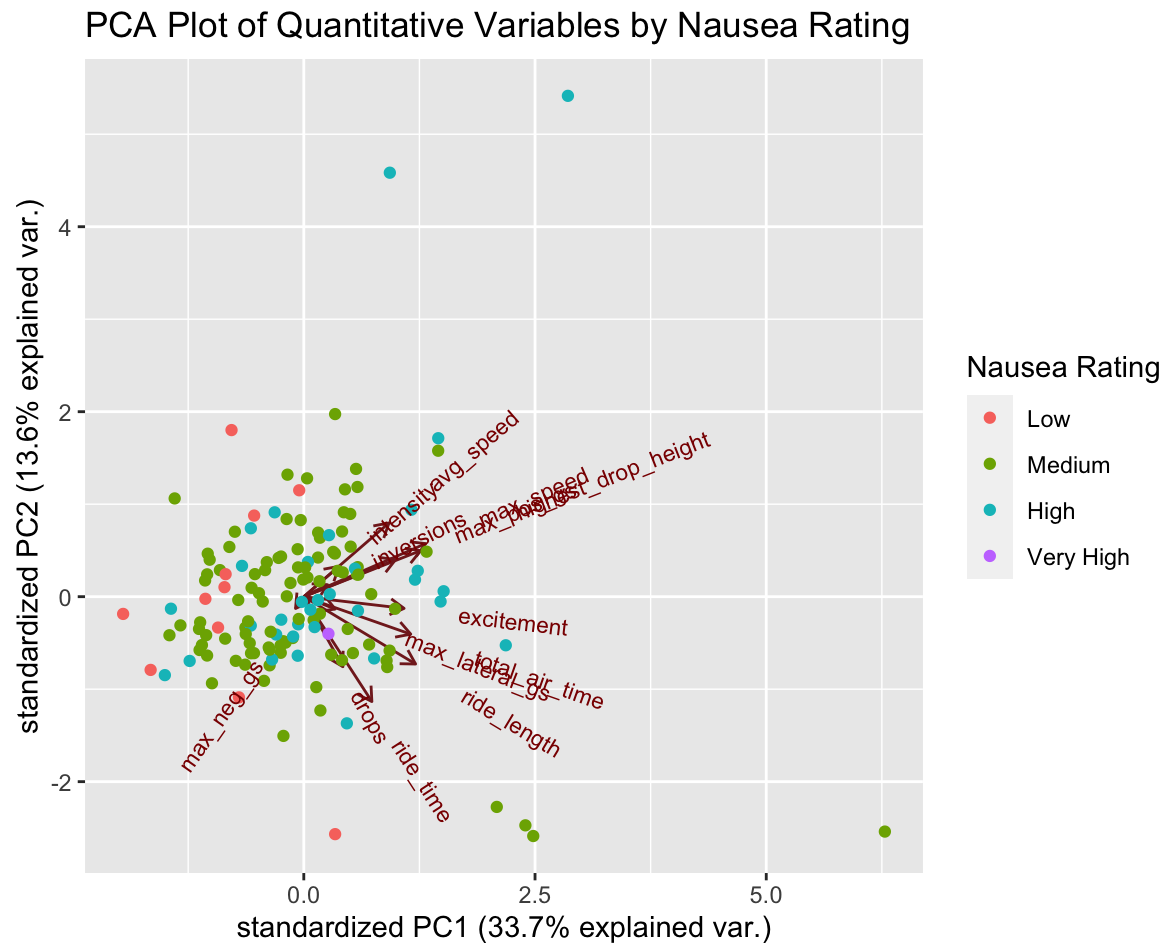
Weak = weak inversions



Here, we plot a faceted scatterplot, and we notice that ranging from different inversion levels, the approximate median nausea value doesn't seem to change too much. Also, a higher nausea level tends to have a higher maximal speed during the ride. Moreover, it seems that with the bigger maximal lateral G force during the ride, the nausea level of riders tends to be higher for roller coasters with either high or weak inversions, but for roller coasters with median inversions, the nausea level is higher for smaller maximal lateral G force.

Note: We converted the number of inversions to a categorical variable by considering weak to be less than 1 inversion, median to be less than 2 inversions, and high to be greater than 2 after looking at the distribution of the inversions variable (omitted for space). We also converted the max lateral G force to a categorical variable by casing on if the value was less than the mean (small) or if it was larger than the mean (large).

## Graph 5



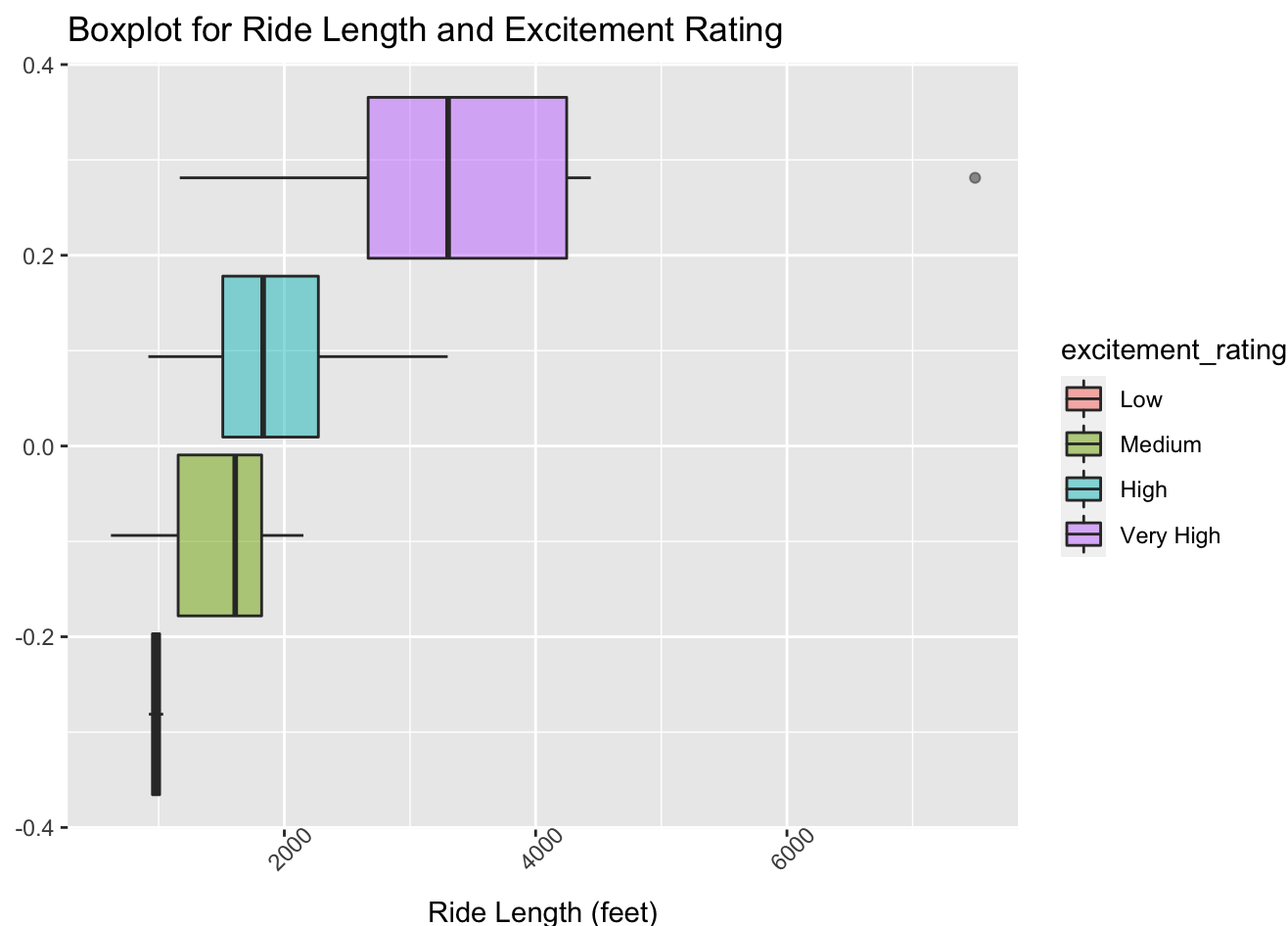
From the biplot, we see that the variables drops, excitement, and max\_lateral\_gs almost point towards the right, thereby signaling that roller coasters with a high first principal component (nausea\_rating) tend to have higher values of these variables. Particularly, we find that drops and ride\_time somewhat point towards the purple point (denoting very high nausea\_rating), further indicating that roller coasters with higher nausea ratings may have longer ride time and higher drops. We note there is only one observation of very high nausea rating, so further analysis with more data points would be needed. Moreover, roller coasters with medium nausea ratings appear to have higher intensity and larger average speed. Finally, we notice that the red points (denoting low nausea rating) tend to be located on the left of the biplot. Since there are almost no variables that point to the left, it seems to be difficult to conclude how low nausea rating relates to the variables.

## Research Question 3: Excitement

We are now exploring the factors that are related to the excitement rating. We will focus mainly on ride length of the roller coaster and also ride time and air time.

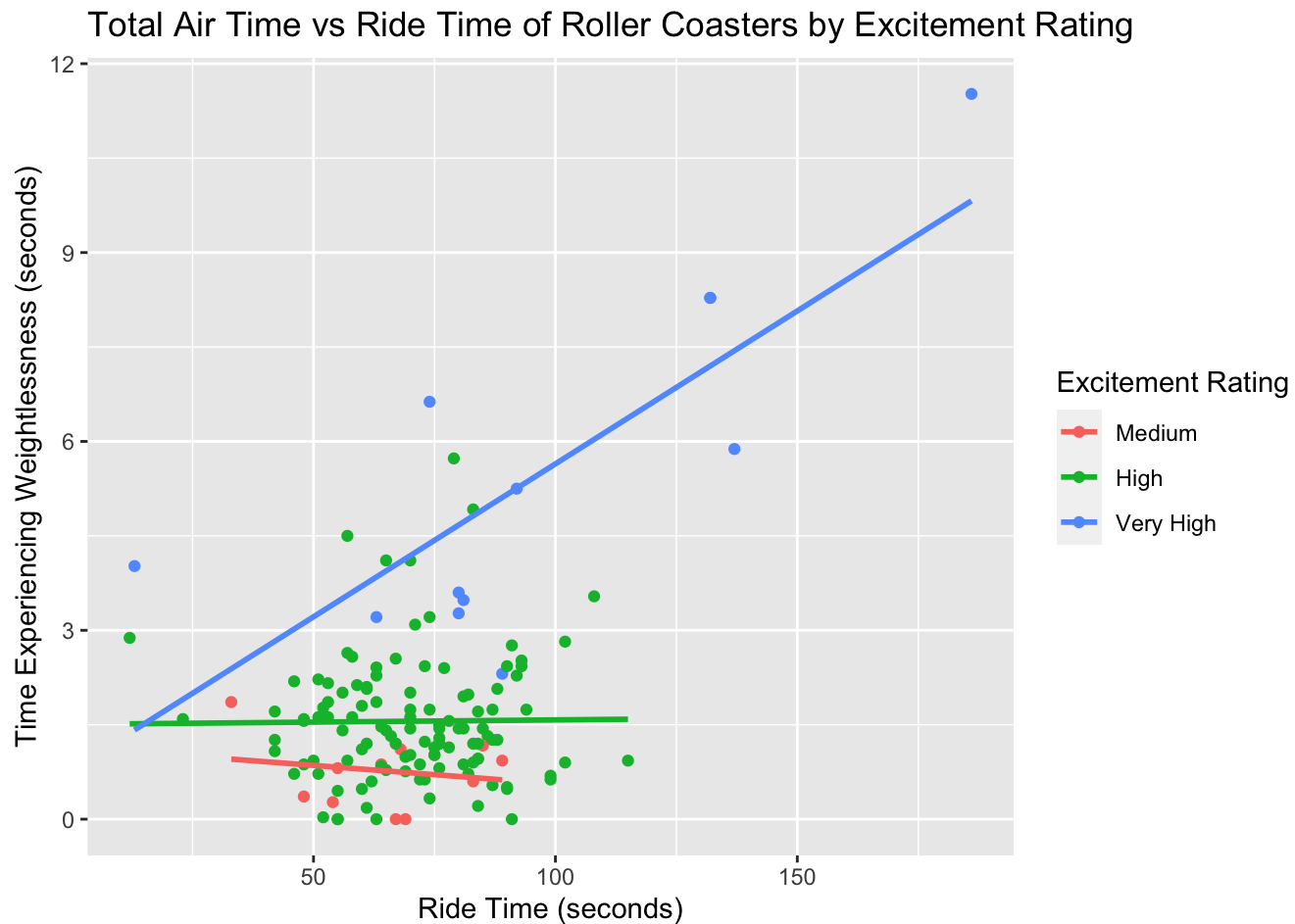
As follows, we are going to discuss the relationship between ride length and excitement rating.

## Graph 6



This graph shows the counts of different ride lengths for different excitement rating levels. The graph shows that for low excitement rating levels, the mean for ride length is approximately 1000 feet. The mean is around 1600 feet for medium level and 1800 feet for high level. For very high level, the mean is much higher than other levels at approximately 3300 feet. Overall, we can see that as the excitement level increases, the median ride length generally increases. Thus, ride length is a potential influencing factor for excitement rating.

## Graph 7



The main takeaway from the scatterplot above is that the relationship between ride time and time experiencing weightlessness differs for the various excitement ratings. We see that an increase in ride time is associated with an increase in mean time experiencing weightlessness for roller coasters with very high excitement rating, barely no change in mean time experiencing weightlessness for roller coasters with high excitement rating, and a decrease in mean time experiencing weightlessness for roller coasters with medium excitement rating. Note that the relationship between ride time and time experiencing weightlessness is very different for roller coasters with very high excitement ratings than those with high or medium excitement ratings. One reason could be because roller coasters with high ride times or high weightlessness times have very high excitement ratings.

Note: There were only 2 data points for low excitement rating, which we removed because the lack of data made it difficult to plot and would not provide a full picture into the trends for low excitement rating roller coasters.

## Conclusion

Overall, there is not much difference between custom designed and pre-designed roller coasters, at least in terms of the quantitative variables we explored in the dataset like ride length. While we did find that custom designed roller coasters increase nausea rating per each increase in excitement rating at a lower rate than pre-designed roller coasters, it's unclear if the difference in the rates is a significant difference.

We were also able to infer that the variables, such as max\_lateral\_gs, max\_speed, and drops in the dataset, appear to strongly predict the nausea or nausea rating. Additionally, we might conclude that the number of inversions don't seem to affect the nausea value.



Furthermore, we concluded that the relationship between ride time and time experiencing weightlessness could impact excitement rating. Specifically, there is a positive association between ride time and time experiencing weightlessness for roller coasters with very high excitement rating, barely any association between ride time and time experiencing weightlessness for roller coasters with high excitement rating, and a negative association between ride time and time experiencing weightlessness for roller coasters with low excitement rating. We also saw that the median ride length increases for each increasing excitement rating.

For future work, since there were so few data points for some levels of the categorical variables excitement rating and nausea rating, it may be beneficial to conduct further analysis with a larger dataset that encompasses more data points for these more “rare” categories of roller coasters. Additionally, because there are many metrics for the roller coasters like max speed and max lateral gs, some future work could also include exploring how these metrics relate to one another and how roller coaster construction could maximize or minimize certain relationships.